Probing command following in patients with disorders of consciousness using a brain–computer interface

Dorothée Lulé, Quentin Noirhomme, Sonja C. Kleih, Camille Chatelle, Sebastian Halder, Athena Demertzis, Marie-Aurélie Bruno, Olivia Gosseries, Audrey Vanhaudenhuyse, Caroline Schnakers, Marie Thonnard, Andrea Soddu, Andrea Kübler, Steven Laureys

Coma Science Group, Cyclotron Research Centre, University and University Hospital of Liège, Liège, Belgium
Institute of Medical Psychology and Behavioural Neurobiology, University of Tübingen, Tübingen, Germany
Department of Psychology I, University of Würzburg, Würzburg, Germany
Department of Neurology, University of Ulm, Ulm, Germany

Objective: To determine if brain–computer interfaces (BCIs) could serve as supportive tools for detecting consciousness in patients with disorders of consciousness by detecting response to command and communication.

Methods: We tested a 4-choice auditory oddball EEG-BCI paradigm on 16 healthy subjects and 18 patients in a vegetative state/unresponsive wakefulness syndrome, in a minimally conscious state (MCS), and in locked-in syndrome (LIS). Subjects were exposed to 4 training trials and 10 –12 questions.

Results: Thirteen healthy subjects and one LIS patient were able to communicate using the BCI. Four of those did not present with a P3. One MCS patient showed command following with the BCI while no behavioral response could be detected at bedside. All other patients did not show any response to command and could not communicate with the BCI.

Conclusion: The present study provides evidence that EEG based BCI can detect command following in patients with altered states of consciousness and functional communication in patients with locked-in syndrome. However, BCI approaches have to be simplified to increase sensitivity.

Significance: For some patients without any clinical sign of consciousness, a BCI might bear the potential to employ a “yes–no” spelling device offering the hope of functional interactive communication.

1. Introduction

Disorders of consciousness (DOC) present a challenging diagnosis and management in clinical routine. Surviving coma, some patients may “awaken” (meaning they open their eyes) but show no clinical sign of awareness of themselves or their environment, i.e., showing only reflex behavior. They are then assumed to be in a “vegetative state” (Jennett and Plum, 1972), recently renamed “unresponsive wakefulness syndrome” (VS/UWS) (Laureys et al., 2010). After some time, if they start to present with non-reflex movement (e.g., visual fixation or pursuit, localization to pain) or to show command following (e.g., to “please squeeze my hand”) they will be considered in a minimally conscious state (MCS) (Bruno et al., 2011b; Giacino et al., 2002). Apparition of reliable communication indicates the emergence from the MCS. Locked-in syndrome patients (LIS) (Plum and Posner, 1972) are fully conscious but are completely paralyzed except for small movements of the eyes or eyelids which may result in confusion with VS/
UWS. Physicians realize that a LIS patient is conscious and able to communicate via eye movements in only 62% of cases (Bruno et al., 2008). Additionally, the time elapsed between the insult and the diagnosis of LIS is about 78 days (Laureys et al., 2005). In spite of what healthy subjects might think, LIS patients can experience a good quality of life (Bruno et al., 2011a; Lule et al., 2009). In the case of a total LIS, an extreme form of LIS with no behavioral signs of consciousness such as oriented eye movements, the rate of mis-diagnosis is likely to be even higher (Schnakers et al., 2009a). Key-stones in diagnosis are the acquisition of voluntary responses such as command following and functional communication which indicates emergence from the VS/UWS (Schnakers et al., 2009c) and the minimally conscious state, respectively. Command following and functional communication also distinguishes LIS from the VS/UWS and MCS patients. In some patients, recovery of consciousness may precede motor recovery. Recent functional neuro-imaging studies based on active tasks provided evidence for awareness in patients diagnosed with VS/UWS or MCS as they presented with clear signs of awareness and volitional control of brain functions detected with functional magnetic resonance imaging (fMRI) (Bardin et al., 2011; Boly et al., 2010; Owen et al., 2006), electroencephalography (EEG) (Schnakers et al., 2008; Goldfine et al., 2011; Cruse et al., 2011) or electromyography (EMG) (Bekinschtein et al., 2008). An active EEG paradigm also enabled correct diagnosis of a total LIS patient (Schnakers et al., 2009b). Brain–computer interfaces (BCI) might permit these patients to show non-motor dependent signs of awareness and in a next step might enable communication (Kübler and Kotchoubey, 2007; Kübler, 2008; Laureys and Boly, 2008).

BCIs can be utilized for communication and interaction with the environment if voluntary muscle activity is lost, i.e., using only the electrical activity of the brain (Wolpaw et al., 2002). A BCI provides subjects with a virtual keyboard, whose keys are pressed by the modulation of brain activity. Each key press constitutes a choice of an item from a set contained in the keyboard, and the subject's choice is indicated through the control of electrical brain activity (Sellers and Donchin, 2006; Sellers et al., 2006). The non-invasive recording of the EEG is the most frequently and feasibly used method in BCI research (Kübler and Kotchoubey, 2007). Present-day BCIs determine the intent of the user from a variety of different electrophysiological signals. These signals include slow cortical potentials, P3 potentials, steady-state evoked potentials, and mu or beta rhythms (Kübler and Müller, 2007; Wolpaw et al., 2002). A specific algorithm translates the extracted features into commands that represent the users' intent. These commands can control effectors to select items such as "yes–no" choices or word spelling for device communication.

To detect command following, we need a signal which can be controlled with a minimum of training and a low workload like P3 or steady-state evoked potentials which are based on automatic brain responses modulated by attention. The visual P3 evoked potentials are presently the most robust BCI enabling accuracy close to 100% in healthy subjects (Bin et al., 2009; Blankertz et al., 2010; Sellers et al., 2006) and P3 paradigms also allow for high accuracy in patients with severe motor paralysis (e.g., see Nijboer et al., 2010; Kleih et al., 2011 for a review). However, DOC patients do not always control eye movements which may cause difficulties in orienting attention to a specific location in the visual field. Therefore, we here chose to focus on a non-visual BCI paradigm based on a four-choice auditory oddball paradigm to provide users with the ability to answer simple questions (Sellers and Donchin, 2006). To date, few studies with non-visualy based P3-BCI exist (e.g., see Klohsa et al., 2009; Furdea et al., 2009; Kübler et al., 2009; Schreuder et al., 2011 for auditory BCIs). Halder et al. developed an auditory BCI based on a three-stimulus paradigm similar to the standard oddball. This BCI has a high viability for fast binary communication (yes–no) independent of vision (Halder et al., 2010). In practice, this paradigm is similar to the manner of communication used by locked-in syndrome (LIS) patients who retain some rudimentary muscle control (e.g., patient may respond "yes" by looking to the right, or raising an eyebrow, and look down for a "no" response). Sellers and Donchin tested an auditory P3-based BCI as a non-muscular communication device in severely physically disabled patients (2006). We here adopted the approach of Sellers and Donchin (2006) of a four-choice auditory stimulus paradigm. In this paradigm based on a P3-BCI four stimuli were presented to the subjects ("yes", "no", "pass", "end") who had to focus on either the target "yes" or "no". The probability of each target event was low (25%) and therefore the targets were likely to yield a detectable P3 (Duncan-Johnson and Donchin, 1977; Johnson and Donchin, 1978). The authors showed that the response remained stable over a period of ten sessions in healthy volunteers as well as in patients with amyotrophic lateral sclerosis (ALS); although accuracy was lower in patients (Sellers and Donchin, 2006).

With this study we aimed at testing to what extent a four-choice auditory P3-based BCI (Furdea et al., 2009; Sellers and Donchin, 2006) could help detecting response to command and functional communication in patients with altered states of consciousness. To the best of our knowledge, this is the first attempt to test a P3-based BCI in a population with DOC.

2. Materials and methods

We studied 18 severely brain damaged patients who had survived a coma. Patients were evaluated repeatedly using the Coma Recovery Scale Revised (CRS-R) (Kalmar and Giacino, 2005). Two patients were in LIS (aged 63 and 29, both female, following brainstem stroke, time post injury 26 and 46 months), 13 were diagnosed with MCS (aged 42 ± 21 years, 9 male, 5 of traumatic etiology, mean time post injury 70 ± 109 months) and three patients were in a VS/UWS (aged 61 ± 17 years, 1 female, 2 with anoxic etiology, time post injury 10 ± 15 months). Additionally, 16 healthy subjects (aged 45 ± 19, 7 males) were included as control group: ten younger aged healthy controls <55 years, mean age 29 ± 6, and six older aged healthy controls >55 years, mean age 66 ± 7 years.

The study was approved by the Ethical Review Board of the Medical Faculty, University of Liège. Each subject was informed about the purpose of the study and signed informed consent prior to participation. Locked-in state patients were required to indicate agreement via their legal representative and via eye-coded communication. For non-communicative patients, the legal representative was informed about the purpose of the study and signed informed consent.

2.1. Procedure

An auditory P3 four choice speller paradigm (Furdea et al., 2009; Sellers and Donchin, 2006) based on the BCI2000 system (Schalk et al., 2004) was used. Auditory stimuli were presented with two separate speakers approximately 1 meter apart, positioned in front and directed toward the participants. All participants took part in one experimental session, which consisted of two experimental runs.

The first run comprised of 4 trials and the second run of 12 trials for healthy subjects and 10 trials for patients (Fig. 1). The number of trials was reduced for patients to decrease duration of the experiment and subsequent risk of fatigue. Users were presented with four stimuli ("yes", "no", "stop", "go") in a random sequence. French ("oui", "non", "stop", "va") and German ("ja", "nein", "stop", "go") versions of the stimuli were used depending of subjects' mother
tongue. Each trial encompassed 15 presentations of four words (60 words in total). The order of presentation was pseudo-randomized (sound duration: 400 ms; inter-stimulus interval: 600 ms, a trial lasting about 1 min). The participants’ task was to count the number of times a target, either “yes” or “no”, was presented.

The first run was intended to test command following and to provide a learning dataset to determine the coefficients of the discriminant function (Furdea et al., 2009; Krusienski et al., 2006). Participants were asked to concentrate on either yes or no determined by the experimenter (two times on “yes” and two times on “no”). Command following was tested post hoc by detecting the presence/absence of a P3 in response to the target. A discriminant function was trained on the data of the first run, and was then used for online classification. The training of the discriminant function took less than a minute on a standard laptop. In the second run, patients and healthy subjects were required to answer 10 or 12 questions, respectively. Questions were of the following kind: “Is your name Quentin?”, “Is your mother’s name Dorothée?”. Feedback of the selected word was provided to the experimenter after each selection in less than a second and was communicated to the patient after each positive trial. Each trial was followed by a short break which could last several minutes with patients. Test runs followed immediately the calibration run due to low attention span of the patients and to test the P3 paradigm in a real clinical environment (i.e., insuring external validity).

2.2. Data acquisition

Stimulus presentation and data collection were controlled by the BCI2000 software (Schalk et al., 2004; http://www.bci2000.org/). The EEG was recorded using an Ag/AgCl electrode cap with 16 channels (F3, Fz, F4, T7, T8, C3, Cz, C4, Cp3, Cp4, P3, Pz, P4, PO7, PO8, and Oz) based on the international 10–20 system (Sharbrough et al., 1991). Each channel was referenced to the right and grounded to the left mastoid. The EEG was amplified using a gUSBamp 16-channel amplifier (g-tec medical engineering GmbH, Schiedlberg, Austria), sampled at 256 Hz, band-pass filtered between 0.01–30 Hz. Data processing, storage, and on-line display of the participants’ EEG were conducted using a Dell Latitude D600 laptop.

2.3. Data analysis

The response to command was tested offline with the data of the first run. The four trials were concatenated (60 target and 180 non-target events) and analyzed with SPM8 (Litvak et al., 2011). After filtering between 0.5 and 20 Hz, the data were epoched between −200 and 1000 ms and then averaged using the “robust average method” (Litvak et al., 2011). Robust average uses statistical distribution of values over trials to eliminate artifacts. This approach is user independent and suppresses artifacts restricted to narrow time and frequency ranges, without rejecting whole trials. For each subject, P3 identification was performed in a two-step process. At each sampling point of each derivation, we calculated the significance of the averaged potential using a permutation test with 1000 permutations (p < 0.05) (Nichols and Holmes, 2002). A P3 wave was detected when a significant positive deflection in the latencies between 250 and 1000 ms around the centro-parietal area (C3, Cz, C4, CP3, CP4, P3, P4, PO7, PO8, and Oz) lasted at least 60 ms (Fischer et al., 2008). Furthermore, all averaged data were visually inspected.

Functional communication was tested with a stepwise linear discriminant analysis method (SWLDA) for online classification of subjects’ responses (Donchin et al., 2000; Farwell and Donchin, 1988; Krusienski et al., 2006). The method, an extension of the Fisher’s Linear Discriminant, is well established as a successful classification method for P3-BCI data. The algorithm seeks an optimal discriminant function by adding spatiotemporal features (amplitude values from particular channel locations and time samples) to a linear equation (Furdea et al., 2009; Krusienski et al., 2006). First run data (training session) was used to build the classifier. The classifier was then immediately used for the online run data in which questions had to be answered. To assess the significance of the results we calculated the significance of observed frequencies (SoF; Kübler and Birbaumer, 2008). The observed frequencies of hits (correctly chosen target) and misses were compared to the expected frequencies given chance performance and tested for significance as follows:

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

where fo is the observed frequency and fe the expected frequency and degree of freedom df = 1. Using 10 and 12 questions for patients and controls, respectively, in a four choice paradigm 2.5 and 3 correct answers can be expected by chance. $\chi^2$ needs to be >3.8 to yield significant observations. This equals 6 or more correct answers (fo) for 10 questions ($\geqslant 60\%$) and 7 or more correct answers for 12 questions ($\geqslant 58.3\%$).

Change in vigilance often occurs in patients with DOC. In case it occurs during a session, it could badly influence the final results. One bad training trial may prevent good training of the classifier. Increasing the pool of training data may partially overcome this problem. Therefore, offline, all training and testing trials were pooled. Trials with artifacts were discarded and a leave-one-out approach was used to classify the data. Then, the significance of the results was calculated with the SoF formula. Significant offline result were admitted as a proof of command following.

3. Results

3.1. Response to command

Eleven healthy controls presented with a P3 following the presentation of the target stimulus (Fig. 2, left). One healthy control

---

had an observable but earlier component peaking at 280 ms. Four healthy controls presented an observable negative deflection which is also often seen as a reaction to P3-BCI stimulation (Allison and Pineda, 2003; Bianchi et al., 2010; Kaufmann et al., 2011). One control subject had no observable deflection. One LIS patient presented with a P3 (Fig. 2, middle). One MCS patient showed a P3 like brain signal on the right hemisphere of the brain but significant only on T8 (Fig. 2, right). No stimulus related response could be detected for any of the other patients.

3.2. Functional communication

Healthy subjects presented a mean correct response rate of 73 ± 23% online; only three of all healthy subjects did not show a significant χ². Of those three, two still achieved an accuracy of 50%. One LIS patient had a significant correct response rate of 60% while the other LIS patient had a response rate of 20% and could not use the BCI for communication. While the classifier was always able to compute discriminant weights on the basis of the training data, no MCS or VS/UWS patient could functionally communicate with the BCI (Table 1).

3.3. Offline classification

In all but two healthy subjects, classification rate could be improved offline. Healthy subjects presented a mean offline correct response rate of 87 ± 29.4%. Two of the three subjects who did not reach significant online results reached significant offline accuracy. Likewise, in the two LIS patients accuracy could be improved offline (see Table 1). MCS patients had inconsistent offline results: Some presented with a decreased accuracy while others improved. Three MCS patients reached an accuracy of 50% or more and one of them obtained significant results (see Table 1). This one MCS patient did not show any command-following at bedside. All VS/UWS patients had a drop in accuracy. The drop in accuracy from online to offline is due to the increased training set used for offline analysis. Increasing the number of samples used for training may increase the accuracy of SWLDA (Blankertz et al., 2011). Therefore, our offline results are less subject to chance.

4. Discussion

The presented auditory P3-based BCI was able to detect offline command following in one MCS patient who never showed bedside clinical signs of command following. Moreover, it allowed us to establish functional communication in one of two locked-in patients. Other functional neuroimaging approaches have been suggested to detect command following in DOC patients with both fMRI (Bardin et al., 2011; Boly et al., 2007; Monti et al., 2010; Owen et al., 2006) and an active EEG paradigms (Schnakers et al., 2008, 2009a; Goldfine et al., 2011; Cruse et al., 2011). The use of fMRI limits the type of patients who can be assessed, excluding patients with ferromagnetic implants or patients with involuntary movement, and limits the possibility of repeated assessment in patients with a level of vigilance fluctuating across the day. In contrast, an EEG based approach can be easily set up at the patient’s bedside, which may render this approach more applicable in clinical routine. A BCI can detect any significant activations evoked by a stimulus independently of their shape or latency (Allison and Pineda, 2003; Bianchi et al., 2010; Kaufmann et al., 2011), which is particularly important for patients with DOC as they may present with event-related responses quite different in shape, latency or topography as compared to healthy subjects.

Four patients in a MCS with clinical command following did not show a detectable response in the BCI. The success of control that is achieved by means of any BCI system still greatly varies between subjects. Training of BCI control might improve performance (e.g., how well a person can concentrate on the target) and it may need several training sessions to improve accuracy and to obtain significant results, specifically in patients with severe brain damage (Kübler and Birbaumer, 2008). The absence of significant results may also be due to the rapid changes in vigilance in DOC patients. Change in vigilance can only be taken into account by multiple recording (which was not possible with the patients of the current study) or by BCIs that monitor arousal and detect states of low vigilance or sleep (which are not yet available).

To cope with the possible issue of fatigue, we tried to keep the experiment as short as possible with only four training trials and a limited number of questions. A single experiment with both training and questions may reduce the influence of change in vigilance but at the cost of a low number of training trials for the classifier. Visual-P3 BCIs, in comparison to the proposed approach, usually use 900 trials (150 targets) to train the classifier (Guger et al., 2009). This number is only matched during the offline analysis. In patients with DOC, the limited number of training trials will always be a challenge. Also, due to time constraints, we did not test the obtained classifier before applying the questions. Thus, the low rate of correct responses could also be due to not optimal calibrated classifiers.
Severely brain damaged patients emerging from MCS may have difficulty giving accurate answers to simple yes–no questions (Nakase-Richardson et al., 2009) and could present aphasia (Majerus et al., 2009). Thus, our questions may have been too difficult to answer or we may simply have posed them in a state when the patients were not conscious.

Suboptimal EEG recording quality due to movement, ocular and respiration artifacts in these challenging populations may also be confounding factors. Further algorithmic developments and improvement of classification could improve BCI control in DOC patients by including automatic artifact detection, single trial classification and data classification for which no training is needed (De Vico Fallani et al., 2011; Hinterberger et al., 2003). Recently proposed classifiers such as shrinkage-RLDA (Blankertz et al., 2011) or Bayesian-LDA (Hoffmann et al., 2008) may prove more suitable for such difficult data. Furthermore, other brain signals such as steady state evoked potentials currently used in BCI should be tested. We here chose a P3-BCI for its robustness and extensive experience within BCI (Kleih et al., 2011), and given that movement imagery has been tested with success (Goldfine et al., 2011; Cruse et al., 2011), and that, although, visual P3 paradigms show higher accuracy than auditory P3 paradigm (Furdea et al., 2009), they require control of eye movement which is not commonly present in patients with DOC.

To summarize, whether the P3 four-choice speller can ever be used for reliable communication in patients with DOC cannot be answered at the present time. Future research needs to further elucidate how well LIS patients, of who we know they are conscious, can control these BCIs and whether BCI command-following in patients with DOC can be established with BCI. Further, it has to be taken into account that for about 20% of potential users the obtained accuracy does not reach criterion level, meaning that BCI control is not accurate enough to control an application (Kübler and Birbaumer, 2008; Blankertz et al., 2010).

Despite the need for further adaptation of BCI to post-coma patients with disorders of consciousness, our results already indicate that the presented EEG-BCI could be applicable in clinical settings to detect signs of consciousness in MCS and LIS patients. Although communication might not be feasible for all patients, BCI might still be the key for the conscious brain locked into a paralyzed body (Kübler and Neumann, 2005; Laureys and Boly, 2008). For some patients without any clinical sign of command-following (in this study one patient with MCS) BCI might bear the potential to re-install simple yes–no communication offering the hope of re-inclusion and thus, autonomy.

### Acknowledgements

S. L. is Senior Research Associate; Q. N., M.-A. B., C. S. and A. V. are Postdoctoral Researcher and C. C. and O. G. are Research Fellow at the Fonds de la Recherche Scientifique (FRS). This work is supported by the European ICT Programme Projects FP7-247919 DECODER, Marie Curie DISCOS ITN, COST, James McDonnell Foundation, Mind Science Foundation, FRS and University and Uni-